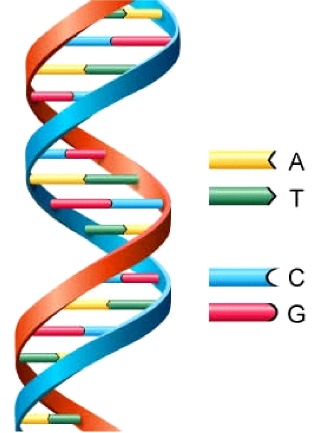
Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Period: \_\_\_\_\_\_

**DNA Statistics Assignment**

Note: Be sure you can locate all of the following formulas on your AP Equation Sheet as this is what will be provided on your exam

**Hypothesis:**

Historically, Erwin Chargaff used the numbers of each of the nitrogenous bases to determine his famous base pairing rule: A-T and C-G. Additionally, scientists are able to compare samples of DNA to determine how genetically similar organisms are to one another. The more DNA they have in common, the closer the relatedness between the organisms. The following is genetic information for three different organisms. We are going to be investigating as to whether there is a significant difference between the amounts of A-T pairings and the amount of C-G pairings. In this example, you can think of the organisms as the “trials” of the experiment to determine if there is a significant difference between A-T pairings and C-G pairings.

1. What is the null hypothesis for this set of data?

*In this case, the null hypothesis would be that there is no significant difference between the amount of A-T pairings and the amount of G-C pairings.*

1. What is the alternate hypothesis for this set of data?

**Statistical Procedure**

The following data is a set of data collected from three separate organisms. Please note that the values for each nitrogenous base are provided in the millions

|  |  |  |
| --- | --- | --- |
|  | **# of Thymine and Adenine Pairs** | **# of Guanine and Cytosine Pairs** |
| Organism #1 | 180 | 90 |
| Organism #2 | 160 | 100 |
| Organism #3 | 143 | 117 |
| Organism #4 | 125 | 125 |
| Organism #5 | 158 | 111 |

1. Use the data in the chart on the previous page to calculate the **mean** for each set of data.

Record your calculated means in the table given below.

|  |  |  |
| --- | --- | --- |
|  | **# of Thymine and Adenine Pairs** | **# of Guanine and Cytosine Pairs** |
| (Mean) |  |  |

1. Use the data in the chart on the previous page to calculate the **standard deviation** for each set of data.

Record your work in row 1 of the table below, and record your calculated standard deviations in row 2 of the table given below.

|  |  |  |
| --- | --- | --- |
|  | **# of Thymine and Adenine Pairs** | **# of Guanine and Cytosine Pairs** |
| Work |  |  |
| s (standard deviation) |  |  |

3. Use the standard deviations you calculated to find the **standard error of the mean (SEM)** for each set of data.

Record your work in row 1 of the table below, and record your calculated SEMs in row 2 of the table given below.

|  |  |  |
| --- | --- | --- |
|  | **# of Thymine and Adenine Pairs** | **# of Guanine and Cytosine Pairs** |
| Work |  |  |
| (SEM) |  |  |

4. Next you will calculate the **95% confidence limit.**

***Formula:*** 95% Cl = Mean ± 2SEM

***Remember..***

The (mean + 2SEM) gives you the **error bar upper limit** (the top of the error bar). The (mean – 2SEM) gives you the **error bar lower limit** (the bottom of the error bar). The error bar provides a range of values around the mean for this sample of data. If we sampled a much larger number of data points (in this case, more organisms), we would be 95% confident that the mean of this larger sample was within the error bar range.

Record your calculated error bar upper limits in row 1 of the table below, and record your calculated error bar upper limits in row 2 of the table below.

|  |  |  |
| --- | --- | --- |
|  | **# of Thymine and Adenine Pairs** | **# of Guanine and Cytosine Pairs** |
| Error Bar Upper Limit |  |  |
| Error Bar Lower Limit |  |  |

5. In the grid given, create a bar graph showing the mean # of A-T pairings vs. C-G pairings. For each set of data, please graph an error bar using the error bar upper limit and error bar lower limit. Make sure to include the following elements of a proper scientific graph.

-A title that includes information about the x and y axis

-X and Y axis labels (with units where applicable)

-An appropriate scale on your X and Y axes



1. Is there overlap in the error bars for the two data sets? What does this mean in regards to our data?